



## Theorems on Positive Data

### *on the Uniqueness of NMF*

Laurberg, Hans; Christensen, Mads Græsbøll; Plumbley, Mark D.; Hansen, Lars Kai; Jensen, Søren Holdt

*Published in:*  
Computational Intelligence and Neuroscience

*DOI (link to publication from Publisher):*  
[10.1155/2008/764206](https://doi.org/10.1155/2008/764206)

*Publication date:*  
2008

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

*Citation for published version (APA):*  
Laurberg, H., Christensen, M. G., Plumbley, M. D., Hansen, L. K., & Jensen, S. H. (2008). Theorems on Positive Data: on the Uniqueness of NMF. *Computational Intelligence and Neuroscience*.  
<https://doi.org/10.1155/2008/764206>

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

#### Take down policy

If you believe that this document breaches copyright please contact us at [vbn@aub.aau.dk](mailto:vbn@aub.aau.dk) providing details, and we will remove access to the work immediately and investigate your claim.

## Research Article

# Theorems on Positive Data: On the Uniqueness of NMF

**Hans Laurberg,<sup>1</sup> Mads Græsbøll Christensen,<sup>1</sup> Mark D. Plumbley,<sup>2</sup>  
Lars Kai Hansen,<sup>3</sup> and Søren Holdt Jensen<sup>1</sup>**

<sup>1</sup> Department of Electronic Systems, Aalborg University, Niels Jernes Vej 12, 9220 Aalborg, Denmark

<sup>2</sup> Department of Electronic Engineering, Queen Mary, University of London, Mile End Road, London E1 4NS, UK

<sup>3</sup> Department of Informatics and Mathematical Modeling, Technical University of Denmark, Richard Petersens Plads, Building 321, 2800 Lyngby, Denmark

Correspondence should be addressed to Hans Laurberg, hla@es.aau.dk

Received 1 November 2007; Accepted 13 March 2008

Recommended by Wenwu Wang

We investigate the conditions for which nonnegative matrix factorization (NMF) is unique and introduce several theorems which can determine whether the decomposition is in fact unique or not. The theorems are illustrated by several examples showing the use of the theorems and their limitations. We have shown that corruption of a unique NMF matrix by additive noise leads to a noisy estimation of the noise-free unique solution. Finally, we use a stochastic view of NMF to analyze which characterization of the underlying model will result in an NMF with small estimation errors.

Copyright © 2008 Hans Laurberg et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. Introduction

Large quantities of positive data occur in research areas such as music analysis, text analysis, image analysis, and probability theory. Before deductive science is applied to large quantities of data, it is often appropriate to reduce data by preprocessing, for example, by matrix rank reduction or by feature extraction. Principal component analysis is an example of such preprocessing. When the original data is nonnegative, it is often desirable to preserve this property in the preprocessing. For example, elements in a power spectrogram, probabilities, and pixel intensities should still be nonnegative after the processing to be meaningful. This has led to the construction of algorithms for rank reduction of matrices and feature extraction generating nonnegative output. Many of the algorithms are related to the nonnegative matrix factorization (NMF) algorithm proposed by Lee and Seung [1, 2]. NMF algorithms factorize a nonnegative matrix  $V \in \mathbb{R}_+^{n \times m}$  or  $R \in \mathbb{R}_+^{n \times m}$  into two nonnegative matrices  $W \in \mathbb{R}_+^{n \times r}$  and  $H \in \mathbb{R}_+^{r \times m}$ :

$$V \approx R = WH. \quad (1)$$

There are no closed-form solutions to the problem of finding  $W$  and  $H$  given a  $V$ , but Lee and Seung [1, 2] proposed

two computationally efficient algorithms for minimizing the difference between  $V$  and  $WH$  for two different error functions. Later, numerous other algorithms have been proposed (see [3]).

An interesting question is whether the NMF of a particular matrix is unique. The importance of this question depends on the particular application of NMF. There can be two different viewpoints when using a model like NMF—either one can believe that the model describes nature and that the variables  $W$  and  $H$  have a physical meaning or one can believe that the model can capture the part of interest even though there is not a one-to-one mapping between the parameters and the model, and the physical system. When using NMF, one can wonder whether  $V$  is a disturbed version of some underlying  $WH$  or whether the data is constructed by another model or, in other words, a ground truth  $W$  and  $H$  does exist. These questions are important in evaluating whether or not it is a problem that there is another NMF solution,  $W'H'$ , to the same data, that is,

$$V \approx R = WH = W'H'. \quad (2)$$

If NMF is used even though the data is not assumed to be generated by (1), it may not be a problem that there are

several other solutions. On the other hand, if one assumes that a ground truth exists, it may be a problem if the model is not detectable, that is, if it is not possible to find  $\mathbf{W}$  and  $\mathbf{H}$  from the data matrix  $\mathbf{V}$ .

The first articles on the subject was two correspondences between Berman and Thomas. In [4] Berman asked for what amounts to a simple characterization of the class of nonnegative matrices  $\mathbf{R}$  for which an NMF exists. As we shall see, the answer by Thomas [5] can be transferred into an NMF uniqueness theorem.

The first article investigating the uniqueness of NMF is Donoho and Stodden [6]. They use convex duality to conclude that in some situations, where the column vectors of  $\mathbf{W}$  “describe parts,” and for that reason are nonoverlapping and thereby orthogonal, the NMF solution is unique.

Simultaneously with the development of NMF, Plumbley [7] worked with nonnegative independent component analysis where one of the problems is to estimate a rotation matrix  $\mathbf{Q}$  from observations on the form  $\mathbf{Q}\mathbf{s}$ , where  $\mathbf{s}$  is a nonnegative vector. In this setup, Plumbley investigates a property for a nonnegative independent and identically distributed (i.i.d.) vector  $\mathbf{s}$  such that  $\mathbf{Q}$  can be estimated. He shows that if the elements in  $\mathbf{s}$  are grounded and a sufficiently large set of observations is used, then  $\mathbf{Q}$  can be estimated. The uniqueness constraint in [7] is a statistical condition of  $\mathbf{s}$ .

The result in [7] is highly relevant to the NMF uniqueness due to the fact that in most cases new NMF solutions will have the forms  $\mathbf{W}\mathbf{Q}$  and  $\mathbf{Q}^{-1}\mathbf{H}$  as described in Section 3. By using Plumbley’s result twice, a restricted uniqueness theorem for NMF can be constructed.

In this paper, we investigate the circumstances under which NMF of an observed nonnegative matrix is unique. We present novel necessary and sufficient conditions for the uniqueness. Several examples illustrating these conditions and their interpretations are given. Additionally, we show that NMF is robust to additive noise. More specifically, we show that it is possible to obtain accurate estimates of  $\mathbf{W}$  and  $\mathbf{H}$  from noisy data when the generating NMF is unique. Lastly, we consider the generating NMF as a stochastic process and show that particular classes of such processes almost surely result in unique NMFs.

This paper is structured as follows. Section 2 introduces the notation, some definitions, and basic results. A precise definition and two characterizations of a unique NMF are given in Section 3. The minimum constraints of  $\mathbf{W}$  and  $\mathbf{H}$  for a unique NMF are investigated in Section 4. Conditions and examples of a unique NMF are given in Section 5. In Section 6, it is shown that in situations where noise is added to a data matrix with a unique NMF, it is possible to bound the error of the estimates of  $\mathbf{W}$  and  $\mathbf{H}$ . A probabilistic view on the uniqueness is considered in Section 7. The implication of the theorems is discussed in Section 8, and Section 9 concludes the paper.

## 2. Fundamentals

We will here introduce convex duality that will be the framework of the paper, but first we shall define the notation to be used. Nonnegative real numbers are denoted as

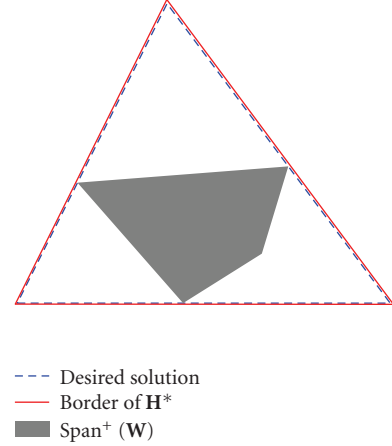


FIGURE 1: A three-dimensional space is scaled such that the vectors are in the hyper plane:  $\{\mathbf{p} \mid [1 \ 1 \ 1]\mathbf{p} = 1\}$ . By the mapping to the hyper plane, a plane in  $\mathbb{R}^3$  is mapped to a line and a simplicial cone is mapped to an area. In the figure, it can be observed that the dashed triangle (desired solution) is the only triangle (third-order simplicial cone) that contains the shaded area (positive span of  $\mathbf{W}$ ) while being within the solid border (the dual of  $\mathbf{H}$ ). The NMF can be concluded to be unique by Theorem 1.

$\mathbb{R}_+$ ,  $\|\cdot\|_F$  denotes the Frobenius norm, and  $\text{span}(\cdot)$  is the space spanned by the set of vectors. Each type of variables has its own font. For instance, a scalar is denoted  $x$ , a column vector is denoted  $\mathbf{x}$ , a row vector is denoted by  $\mathbf{x}$ , a matrix is denoted by  $\mathbf{X}$ , a set is denoted by  $\mathcal{X}$ , and a random variable is denoted by  $\mathcal{X}$ . Moreover,  $\mathbf{x}_i^j$  is the  $i$ th index of the vector  $\mathbf{x}^j$ . When a condition for a set is used to describe a matrix, it is referring to the set of column vectors in the matrix. The NMF is symmetric in  $\mathbf{W}^T$  and  $\mathbf{H}$ , so the theorems for one matrix may also be used for the other.

In the paper, we make a geometric interpretation of the NMF similar to that used in both [5, 6]. For that, we need the following definitions.

**Definition 1.** The *positive span* is given by  $\text{span}^+(\mathbf{b}^1, \dots, \mathbf{b}^d) = \{\mathbf{v} = \sum_i \mathbf{b}^i \mathbf{a}_i \mid \mathbf{a} \in \mathbb{R}_+^d\}$ .

In some literature, the positive span is called the conical hull.

**Definition 2.** A set  $\mathcal{A}$  is called a *simplicial cone* if there is a set  $\mathcal{B}$  such that  $\mathcal{A} = \text{span}^+(\mathcal{B})$ . The *order* of a simplicial cone  $\mathcal{A}$  is the minimum number of elements in  $\mathcal{B}$ .

**Definition 3.** The *dual* to a set  $\mathcal{A}$ , denoted  $\mathcal{A}^*$ , is given by  $\mathcal{A}^* = \{\mathbf{v} \mid \mathbf{v}^T \mathbf{a} \geq 0 \text{ for all } \mathbf{a} \in \mathcal{A}\}$ .

The following lemma is easy to prove and will be used subsequently. For a more general introduction to convex duality, see [8].

**Lemma 1.** (a) If  $\mathcal{X} = \text{span}^+(\mathbf{b}^1, \dots, \mathbf{b}^d)$ , then  $\mathbf{y} \in \mathcal{X}^*$  if and only if  $\mathbf{y}^T \mathbf{b}^n \geq 0$  for all  $n$ .

(b) If  $\mathcal{X} = \text{span}^+(\mathbf{B}^T)$  and  $\mathbf{B}^T = [\mathbf{b}^1, \dots, \mathbf{b}^d]$  is invertible, then  $\mathcal{X}^* = \text{span}^+(\mathbf{B}^{-1})$ .

- (c) If  $\mathcal{Y} \subseteq \mathcal{X}$ , then  $\mathcal{X}^* \subseteq \mathcal{Y}^*$ .  
 (d) If  $\mathcal{Y}$  and  $\mathcal{X}$  are closed simplicial cones and  $\mathcal{Y} \subset \mathcal{X}$ , then  $\mathcal{X}^* \subset \mathcal{Y}^*$ .

### 3. Dual Space and the NMF

In this section, our definition of unique NMF and some general conditions for unique NMF are given. As a starting point, let us assume that both  $\mathbf{W}$  and  $\mathbf{H}$  have full rank, that is,  $r = \text{rank}(\mathbf{R})$ .

Let  $\mathbf{W}'$  and  $\mathbf{H}'$  be any matrices that fulfil,  $\mathbf{R} = \mathbf{W}\mathbf{H} = \mathbf{W}'\mathbf{H}'$ . Then,  $\text{span}(\mathbf{W}) = \text{span}(\mathbf{R}) = \text{span}(\mathbf{W}')$ . The column vectors of  $\mathbf{W}$  and  $\mathbf{W}'$  are therefore both bases for the same space and as a result there exists a basis shift matrix  $\mathbf{Q} \in \mathbb{R}^{r \times r}$  such that  $\mathbf{W}' = \mathbf{W}\mathbf{Q}$ . It follows that  $\mathbf{H}' = \mathbf{Q}^{-1}\mathbf{H}$ . Therefore, all NMF solutions where  $r = \text{rank}(\mathbf{R})$ , are of the form  $\mathbf{R} = \mathbf{W}\mathbf{Q}\mathbf{Q}^{-1}\mathbf{H}$ . In these situations, the ambiguity of the NMF is the  $\mathbf{Q}$  matrix. Note that if  $r > \text{rank}(\mathbf{R})$  the above arguments are not valid because  $\text{rank}(\mathbf{W})$  can differ from  $\text{rank}(\mathbf{W}')$  and thereby  $\text{span}(\mathbf{W}) \neq \text{span}(\mathbf{W}')$ .

*Example 1.* The following is an example of an  $\mathbb{R}_+^{4 \times 4}$  matrix of rank 3, where there are two NMF solutions but no  $\mathbf{Q}$  matrix to connect the solutions

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} = \mathbf{R} = \underbrace{\mathbf{R}}_{\mathbf{W}} \underbrace{\mathbf{I}}_{\mathbf{H}} = \underbrace{\mathbf{I}}_{\mathbf{W}'} \underbrace{\mathbf{R}}_{\mathbf{H}'} \quad (3)$$

We mention in passing that Thomas [5] uses this matrix to illustrate a related problem. This completes the example.

**Lemma 2** (Minc [9, Lemma 1.1]). *The inverse of a non-negative matrix is nonnegative if and only if it is a scaled permutation.*

Lemma 2 shows that all NMF solutions on the forms  $\mathbf{W}\mathbf{Q}$  and  $\mathbf{Q}^{-1}\mathbf{H}$ , where  $\mathbf{Q}$  is a scaled permutation, are valid, and thereby that NMF only can be unique up to a permutation and scaling. This leads to the following definition of unique NMF in this paper.

*Definition 4.* A matrix  $\mathbf{R}$  has a *unique NMF* if the ambiguity is a permutation and a scaling of the columns in  $\mathbf{W}$  and rows in  $\mathbf{H}$ .

The scaling and permutation ambiguity in the uniqueness definition is a well-known ambiguity that occurs in many blind source separation problems. With this definition of unique NMF, it is possible to make the following two characterizations of the unique NMF.

**Theorem 1.** *If  $r = \text{rank}(\mathbf{R})$ , an NMF is unique if and only if the positive orthant is the only  $r$ -order simplicial cone  $\mathcal{Q}$  such that  $\text{span}^+(\mathbf{W}^T) \subseteq \mathcal{Q} \subseteq \text{span}^+(\mathbf{H})^*$ .*

*Proof.* The proof follows the analysis of the  $\mathbf{Q}$  matrix above in combination with Lemma 1(b). The theorem can also be proved by following the steps of the proof in [5].  $\square$

**Theorem 2** (see [6]). *The NMF is unique if and only if there is only one  $r$ -order simplicial cone  $\mathcal{Q}$  such that  $\text{span}^+(\mathbf{R}) \subseteq \mathcal{Q} \subseteq \mathcal{P}$ , where  $\mathcal{P}$  is the positive orthant.*

*Proof.* The proof follows directly from the definitions.  $\square$

The first characterization is inspired by [5] and the second characterization is implicit introduced in [6]. Note that the two characterizations of the unique NMF analyze the problem from two different viewpoints. Theorem 1 takes a known  $\mathbf{W}$  and  $\mathbf{H}$  pair as starting point and looks at the solution from the “inside,” that is, the  $r$ -dimensional space of row vectors in  $\mathbf{W}$  and column vectors in  $\mathbf{H}$ . Theorem 2 looks at the problem from the “outside,” that is, the  $n$ -dimensional column space of  $\mathbf{R}$ .

### 4. Matrix Conditions

If  $\mathbf{R} = \mathbf{W}\mathbf{H}$  is unique, then both  $\mathbf{W}$  and  $\mathbf{H}$  have to be unique, respectively, that is, there is only one NMF of  $\mathbf{W}$  and one of  $\mathbf{H}$ , namely,  $\mathbf{W} = \mathbf{W}\mathbf{I}$  and  $\mathbf{H} = \mathbf{I}\mathbf{H}$ . In this section, a necessary condition for  $\mathbf{W}$  and  $\mathbf{H}$  is given and a sufficient condition is shown.

The following definition will be shown to be a necessary condition for both the set of row vectors in  $\mathbf{W}$  and column vectors in  $\mathbf{H}$ .

*Definition 5.* A set  $\mathcal{S}$  of vectors in  $\mathbb{R}_+^d$  is called *boundary close* if for all  $j \neq i$  and  $k > 0$  there is an element  $\mathbf{s} \in \mathcal{S}$  such that

$$\mathbf{s}_j < k\mathbf{s}_i. \quad (4)$$

In the case of closed sets, the boundary close condition is that  $\mathbf{s}_j = 0$  and  $\mathbf{s}_i \neq 0$ . In this section, the sets will be finite (and therefore closed), but in Section 7 the general definition above is needed.

**Theorem 3.** *The set of row vectors in  $\mathbf{W}$  has to be boundary close for the corresponding NMF to be unique.*

*Proof.* If the set of row vectors in  $\mathbf{W}$  are not boundary close, there exist indexes  $j \neq i$  and  $k > 0$  such that the  $j$ th element is always more than  $k$  times larger than the  $i$ th element in the row vectors in  $\mathbf{W}$ . Let  $\mathcal{Q} = \text{span}^+(\mathbf{q}^1, \dots, \mathbf{q}^r)$ , where

$$\mathbf{q}^n = \begin{cases} \mathbf{e}^i + k\mathbf{e}^j & \text{if } n = i, \\ \mathbf{e}^n & \text{otherwise,} \end{cases} \quad (5)$$

and  $\mathbf{e}^n$  denotes the  $n$ th standard basis vector. This set fulfils the condition  $\text{span}^+(\mathbf{W}^T) \subseteq \mathcal{Q} \subset \mathcal{P}$  and we therefore, using Theorem 1, conclude that the NMF cannot be unique.  $\square$

That not only the row vectors of  $\mathbf{W}$  with small elements determine the uniqueness can be seen from the following example.

*Example 2.* The following is an example where  $\bar{\mathbf{W}}$  is not unique but  $\mathbf{W} = \begin{bmatrix} \bar{\mathbf{W}} \\ 1 \end{bmatrix}$  is.

Let

$$\bar{\mathbf{W}} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}. \quad (6)$$

Here  $\bar{\mathbf{W}}$  is boundary close but not unique since  $\bar{\mathbf{W}} = \bar{\mathbf{W}}\mathbf{I} = \mathbf{I}\bar{\mathbf{W}}$ . The uniqueness of  $\mathbf{W} = \begin{bmatrix} \bar{\mathbf{W}} \\ 1 \end{bmatrix}$  can be verified by plotting the matrix as shown in Figure 1, and observe that the conditions of Theorem 1 are fulfilled. This completes the example.

In three dimensions, as in Example 2, it is easy to investigate whether a boundary close  $\mathbf{W}$  is unique—if  $\mathbf{W} = \mathbf{W}'\mathbf{H}'$ , then  $\mathbf{H}'$  can only have two types of structure: either the trivial (desired) solution where  $\mathbf{H}' = \mathbf{I}$  or a solution where only the diagonal of  $\mathbf{H}'$  is zero. In higher dimensions, the number of combinations of nontrivial solutions increases and it becomes more complicated to investigate all possible nontrivial structures. For example, if  $\bar{\mathbf{W}}$  is the matrix from Example 2, then the matrix

$$\tilde{\mathbf{W}} = \begin{bmatrix} \bar{\mathbf{W}} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{W}} \end{bmatrix} \quad (7)$$

is boundary close and can be decomposed in several ways, for example,

$$\begin{aligned} \tilde{\mathbf{W}} &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{W}} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{W}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \\ &= \begin{bmatrix} \bar{\mathbf{W}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{W}} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{W}} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{W}} \end{bmatrix}. \end{aligned} \quad (8)$$

Instead of seeking necessary and sufficient conditions for a unique  $\mathbf{W}$ , a sufficient condition not much stronger than the necessary is given. In this sufficient condition, we only focus on the row vectors of  $\mathbf{W}$  with a zero (or very small) element.

**Definition 6.** A set of vectors  $\mathcal{S}$  in  $\mathbb{R}_+^d$  is called *strongly boundary close* if it is boundary close, and there exists a  $z > 0$  and a numbering of the elements in the vectors such that for all  $k > 0$  and  $n \in \{1, \dots, d-1\}$  there are  $d-n$  vectors from  $\mathcal{S}$ ,  $\{\mathbf{s}^1, \dots, \mathbf{s}^{d-n}\}$  that fulfil the following:

- (1)  $\mathbf{s}_n^j < k \sum_{i>n} \mathbf{s}_i^j$  for all  $j$ ; and
- (2)  $\kappa_2([\mathbf{b}^1, \dots, \mathbf{b}^{d-n}]) \leq z$ , where  $\kappa_2(\cdot)$  is the “condition number” of the matrix defined as the ratio between the largest and smallest singular values [10, page 81],  $\mathbf{b}^j = \mathbf{P}_n \mathbf{s}^j$  and  $\mathbf{P}_n \in \mathbb{R}^{d-n \times d}$  is a projection matrix that picks the  $d-n$  last element of a vector in  $\mathbb{R}^d$ .

**Theorem 4.** If  $\text{span}^+(\mathbf{W}^T)$  is strongly boundary close, then  $\mathbf{W}$  is unique.

The proof is quite technical and is therefore given in the Appendix. The most important thing to notice is that the necessary condition in Theorem 3 and the sufficient conditions in Theorem 4 are very similar. The first item in the

strongly boundary close definition states that there have to be several vectors with small value. The second item ensures that the vectors with small value are linear independent in the last elements.

## 5. Uniqueness of $\mathbf{R}$

In this section, a condition for unique  $\mathbf{V}$  is analyzed. First, Example 3 is used to investigate when a strongly boundary close  $\mathbf{W}$  and  $\mathbf{H}$  pair is unique. The section ends with a constraint for  $\mathbf{W}$  and  $\mathbf{H}$  that results in a unique NMF.

**Example 3.** This is an investigation of uniqueness of  $\mathbf{R}$  when  $\mathbf{W}$  and  $\mathbf{H}$  are given as

$$\mathbf{H} = \begin{bmatrix} \alpha & 1 & 1 & \alpha & 0 & 0 \\ 1 & \alpha & 0 & 0 & \alpha & 1 \\ 0 & 0 & \alpha & 1 & 1 & \alpha \end{bmatrix}, \quad (9)$$

$$\mathbf{W} = \mathbf{H}^T,$$

where  $0 < \alpha < 1$ . Both  $\mathbf{W}$  and  $\mathbf{H}$  are strongly boundary close and the  $z$  parameter can be calculated as

$$\begin{aligned} z &= \kappa_2([\mathbf{b}^1, \dots, \mathbf{b}^{d-n}]) \\ &= \kappa_2\left(\begin{bmatrix} \alpha & 1 \\ 1 & \alpha \end{bmatrix}\right) = \frac{1+\alpha}{1-\alpha}. \end{aligned} \quad (10)$$

The equation above shows that small  $\alpha$  will result in a  $z$  close to one and an  $\alpha$  close to one results in a large  $z$ . In Figure 2, the matrix  $\mathbf{R} = \mathbf{WH}$  is plotted for  $\alpha \in \{0.1, 0.3, 0.5, 0.7\}$ . The dashed line is the desired solution and is repeated in all figures. It is seen that the shaded area  $\text{span}^+(\mathbf{W}^T)$  is decreasing when  $\alpha$  increases, and that the solid border  $\text{span}^+(\mathbf{H})^*$  increases when  $\alpha$  increases. For all  $\alpha$ -values, both the shaded area and the solid border intersect with the dashed triangle. Therefore, it is not possible to get another solution by simply increasing/decreasing the desired solution. The figure shows that the NMF is unique for  $\alpha \in \{0.1, 0.3\}$  and not unique for  $\alpha \in \{0.5, 0.7\}$  where the alternative solution is shown by a dotted line. That the NMF is not unique for  $\alpha \in \{0.5, 0.7\}$  can also be verified by selecting the  $\mathbf{Q}$  to be the symmetric orthonormal matrix

$$\mathbf{Q} = \mathbf{Q}^T = \mathbf{Q}^{-1} = \frac{1}{3} \begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix}, \quad (11)$$

and see that both  $\mathbf{WQ}$  and  $\mathbf{Q}^{-1}\mathbf{H}$  are nonnegative. If  $\alpha = 0.3$ , then the matrix  $\mathbf{R}$  is given by

$$\mathbf{R} = \frac{1}{100} \begin{bmatrix} 109 & 60 & 30 & 9 & 30 & 100 \\ 60 & 109 & 100 & 30 & 9 & 30 \\ 30 & 100 & 109 & 60 & 30 & 9 \\ 9 & 30 & 60 & 109 & 100 & 30 \\ 30 & 9 & 30 & 100 & 109 & 60 \\ 100 & 30 & 9 & 30 & 60 & 109 \end{bmatrix}. \quad (12)$$

This shows that  $\mathbf{R}$  needs no zeros for the NMF to be unique. This completes the example.



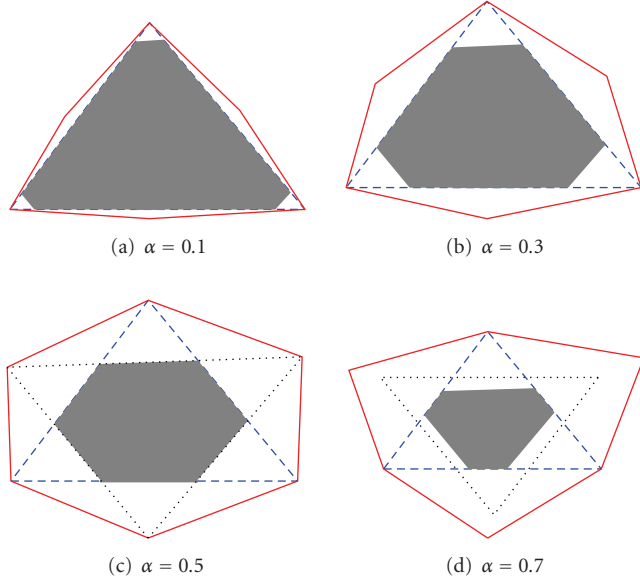


FIGURE 2: The figure shows data constructed as in Example 3 and plotted in the same manner as in Figure 1, that is, the dashed triangle is the desired solution, the solid line is the border of the dual of  $\mathbf{H}$ , and the shaded area is the positive span of  $\mathbf{W}$ . It can be seen that the NMF is unique when  $\alpha$  equals 0.1 or 0.3 but not when  $\alpha$  equals 0.5 or 0.7. In the cases where the NMF is not unique, an alternative solution is shown with a dotted line.

In the example above,  $\mathbf{W}$  equals  $\mathbf{H}^T$  and thereby fulfils the same constraints. In many applications, the meaning of  $\mathbf{W}$  and  $\mathbf{H}$  differs, for example, in music analysis where the column vectors of  $\mathbf{W}$  are spectra of notes and  $\mathbf{H}$  is a note activity matrix [11].

Next, it is investigated how to make an asymmetric uniqueness constraint.

**Definition 7.** A set of vectors in  $\mathbb{R}^d$  is called *sufficiently spread* if for all  $j$  and  $k > 0$ , there is an element  $\mathbf{s} \in \mathcal{S}$  such that

$$\mathbf{s}_j > k \sum_{i \neq j} \mathbf{s}_i. \quad (13)$$

Note that in the definition for sufficiently spread set the  $j$ th element is larger than the sum in contrast to the strongly boundary close definition where the  $j$ th element is smaller than the sum.

**Lemma 3.** The dual space of a sufficiently spread set is the positive orthant.

*Proof.* A sufficiently spread set is nonnegative and the positive orthant is therefore part of the dual set for any sufficiently spread set. Let  $\mathbf{b}$  be a vector with a negative element in the  $j$ th element and select

$$k = \frac{\sum_{i \neq j} |\mathbf{b}_i|}{-\mathbf{b}_j}. \quad (14)$$

In any sufficiently spread set, an  $\mathbf{s}$  exists, such that  $\mathbf{s}_j > k \sum_{i \neq j} \mathbf{s}_i$  and therefore

$$\begin{aligned} \mathbf{s}^T \mathbf{b} &= \mathbf{s}_j \mathbf{b}_j + \sum_{i \neq j} \mathbf{s}_i \mathbf{b}_i \leq \mathbf{s}_j \mathbf{b}_j + \left( \sum_{i \neq j} \mathbf{s}_i \right) \left( \sum_{i \neq j} |\mathbf{b}_i| \right) \\ &= -\mathbf{b}_j \left( -\mathbf{s}_j + k \sum_{i \neq j} \mathbf{s}_i \right) < 0. \end{aligned} \quad (15)$$

The  $\mathbf{b}$  is therefore not in the dual to any sufficiently spread set.  $\square$

In the case of finite sets, the sufficiently spread condition is the same as the requirement for a scaled version of all the standard basis vectors to be part of the sufficiently spread set. It is easy to verify that a sufficiently spread set also is strongly boundary close and that the  $z$  parameter is one.

**Theorem 5.** If a pair  $[\mathbf{W}^T, \mathbf{H}]$  is sufficiently spread and strongly boundary close, then the NMF of  $\mathbf{R} = \mathbf{WH}$  is unique.

*Proof.* Lemma 3 states that the dual set of a sufficiently spread set is the positive orthant,

$$\text{span}^+(\mathbf{H})^* = \mathcal{P} = \text{span}^+(\mathbf{I})^*. \quad (16)$$

Theorem 4 states that  $\mathbf{WI}$  is unique and by using (16) and Theorem 1 we conclude that  $\mathbf{R} = \mathbf{WH}$  is unique.  $\square$

Theorem 5 is a stronger version of the results of Donoho and Stodden [6, Theorem 1]. Theorem 1 in [6] also assumes that  $\mathbf{H}$  is sufficiently spread, but the condition for  $\mathbf{W}^T$  is stronger than the strongly boundary close assumption.

## 6. Perturbation Analysis

In the previous sections, we have analyzed situations with a unique solution. In this section, it is shown that in some situations the nonuniqueness can be seen as estimation noise on  $\mathbf{W}$  and  $\mathbf{H}$ . The error function that describes how close an estimated  $[\mathbf{W}', \mathbf{H}']$  pair is to the true  $[\mathbf{W}, \mathbf{H}]$  pair is

$$J_{(\mathbf{W}, \mathbf{H})}(\mathbf{W}', \mathbf{H}') = \min_{\mathbf{P}, \mathbf{D}} (\|\mathbf{W} - \mathbf{W}'(\mathbf{DP})\|_F + \|\mathbf{H} - (\mathbf{DP})^{-1} \mathbf{H}'\|_F), \quad (17)$$

where  $\mathbf{P}$  is a permutation matrix and  $\mathbf{D}$  is a diagonal matrix.

**Theorem 6.** Let  $\mathbf{R} = \mathbf{WH}$  be a unique NMF. Given some  $\epsilon > 0$ , there exists a  $\delta > 0$  such that any nonnegative  $\mathbf{V} = \mathbf{R} + \mathbf{N}$ , where  $\|\mathbf{N}\|_F < \delta$  fulfils

$$J_{(\mathbf{W}, \mathbf{H})}(\mathbf{W}', \mathbf{H}') < \epsilon, \quad (18)$$

where

$$[\mathbf{W}', \mathbf{H}'] = \arg \min_{\substack{\mathbf{W}' \in \mathbb{R}_+^{n \times r}, \mathbf{H}' \in \mathbb{R}_+^{r \times m}} \|\mathbf{V} - \mathbf{W}' \mathbf{H}'\|_F. \quad (19)$$

The proof is given in the appendix. The theorem states that if the observation is corrupted by additive noise, then it will result in noisy estimation of  $\mathbf{W}$  and  $\mathbf{H}$ . Moreover,

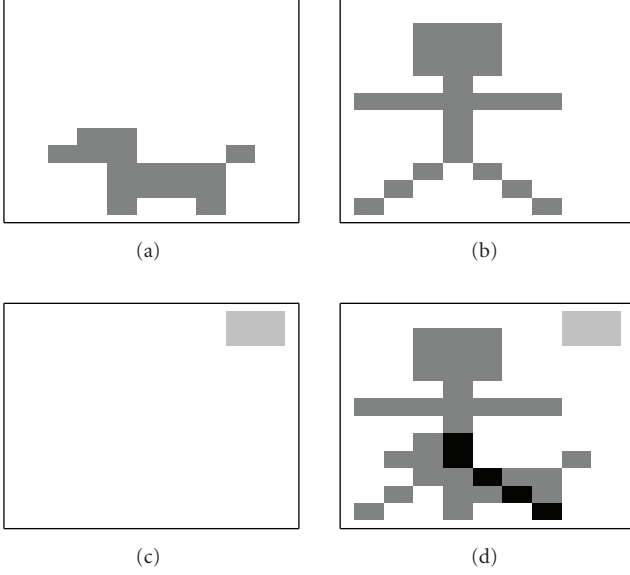


FIGURE 3: The three basis pictures: (a) a dog, (b) a man, and (c) the sun from Example 4, individually and summed in (d).

Theorem 6 shows that if the noise is small, it will result in small estimation errors. In this section, the Frobenius norm is used in (17) and (19) to make Theorem 6 concrete. Theorem 6 is also valid with the same proof if any continuous metric is used instead of the Frobenius norm in those equations.

*Example 4.* This example investigates the connection between the additive noise in  $\mathbf{V}$  and the estimation error on  $\mathbf{W}$  and  $\mathbf{H}$ . The column vectors in  $\mathbf{W}$  are basis pictures of a man, a dog, and the sun as shown in Figures 3(a), 3(b), and 3(c). In Figure 3(d), the sum of the three basis pictures is shown. The matrix  $\mathbf{H}$  is the set of all combinations of the pictures, that is,

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix}. \quad (20)$$

Theorem 5 can be used to conclude that the NMF of  $\mathbf{R} = \mathbf{WH}$  is unique because both  $\mathbf{W}^T$  and  $\mathbf{H}$  are sufficiently spread and thereby also strongly boundary close.

In the example, two different noise matrices,  $\mathbf{N}_N$  and  $\mathbf{N}_M$ , are used. The  $\mathbf{N}_N$  matrix models noisy observation and has elements that are random uniform i.i.d. The  $\mathbf{N}_M$  matrix contains elements that are minus one in the positions where  $\mathbf{R}$  has elements that are two and zero elsewhere, that is,  $\mathbf{N}_M$  is minus one in the positions where the dog and the man are overlapping. In this case, the error matrix  $\mathbf{N}_M$  simulates a model mismatch that occurs in the following two types of real-world data. If the data set is composed of pictures, the basis pictures will be overlapping and a pixel in  $\mathbf{V}$  will consist of one basis picture and not a mixture of the overlapping pictures. If the data is a set of amplitude spectra, the true model is an addition of complex values and not an addition of the amplitudes.

The estimation error of the factorization  $J_{(\mathbf{W}, \mathbf{H})}(\mathbf{W}', \mathbf{H}')$  is plotted in Figure 4 when the norm of the error matrix is  $\mu$ , that is,  $\mathbf{V} = \mathbf{WH} + (\mathbf{N}/\|\mathbf{N}\|_F)\mu$ . An estimate of the  $[\mathbf{W}', \mathbf{H}']$  pair is calculated by using the iterative algorithm for Frobenius norm minimized by Lee and Seung [2]. The algorithm is run for 500 iterations and is started from 100 different positions. The decomposition that minimizes  $\|\mathbf{V} - \mathbf{W}'\mathbf{H}'\|_F$  is chosen, and  $J_{(\mathbf{W}, \mathbf{H})}(\mathbf{W}', \mathbf{H}')$  is calculated numerically. Figure 4 shows that when the added error is small, it is possible to estimate the underlying parameters. When the norm of added noise matrix increases, the behavior of the two noise matrices,  $\mathbf{N}_N$  and  $\mathbf{N}_M$ , differ. For  $\mathbf{N}_N$ , the error of the estimate increases slowly with the norm of the added matrix while the estimation error for  $\mathbf{N}_M$  increases dramatically when the norm is larger than 2.5. In the simulation, we have made the following observation that can explain the difference in the performance of the two types of noise. When  $\mathbf{N}_N$  is used, the basis pictures remain noisy versions of the man, the dog, and the sun. When  $\mathbf{N}_M$  is used and the norm is larger than 2.5, the basis pictures are the man excluding the overlap, the dog excluding the overlap, and the overlap of man and dog. Another way to describe the difference is that the rank of  $\mathbf{N}_M$  is one and the disturbance is in one dimension, where  $\mathbf{N}_N$  is full rank and the disturbance is in many dimensions. This completes the example.

**Corollary 1.** Let  $\mathbf{R} = \mathbf{WH}$  be a unique NMF and  $\mathbf{V} = \widetilde{\mathbf{W}}\widetilde{\mathbf{H}}$ , where  $\widetilde{\mathbf{W}} = \mathbf{W} + \mathbf{N}_W$  and  $\widetilde{\mathbf{H}} = \mathbf{H} + \mathbf{N}_H$ . Given  $\mathbf{R}$  and  $\epsilon > 0$  there exists a  $\delta > 0$  such that if the largest absolute value of both  $\mathbf{N}_W$  and  $\mathbf{N}_H$  is smaller than  $\delta$ , then

$$J_{(\widetilde{\mathbf{W}}, \widetilde{\mathbf{H}})}(\mathbf{W}', \mathbf{H}') < \epsilon, \quad (21)$$

where  $\mathbf{W}'$ ,  $\mathbf{H}'$  are any NMF of  $\mathbf{V}$ .

*Proof.* This follows directly from Theorem 6.  $\square$

The corollary can be used in situations where there are small elements in  $\mathbf{W}$  and  $\mathbf{H}$  but no (or not enough) zero elements—as in the following example.

*Example 5.* Let  $\mathbf{R} = \mathbf{WH}$ , where  $\mathbf{W}$ ,  $\mathbf{H}$  is generated as in Example 3. Let all elements in both  $\mathbf{N}_W$  and  $\mathbf{N}_H$  be equal to  $\eta$ . In Figure 5,  $\mathbf{V}$  is plotted when  $\alpha = 0.3$  and  $\eta = \{0.01, 0.05, 0.10, 0.15\}$ . In this example, neither the shaded area nor the solid border intersect with the desired solution. Therefore, it is possible to get other solutions by simply increasing/decreasing the desired solution. For  $\eta = \{0.01, 0.05\}$ , the corners of the solutions are close to the corners of the desired solution. When  $\eta = 0.1$ , the corners can be placed mostly on the solid border and still form a triangle that contains the shaded area. When  $\eta = 0.15$ , the corners can be anywhere on the solid border. This completes the example.

## 7. Probability and Uniqueness

In this section, the row vectors of  $\mathbf{W}$  and the column of  $\mathbf{H}$  are seen as results of two random variables. Characteristics of the

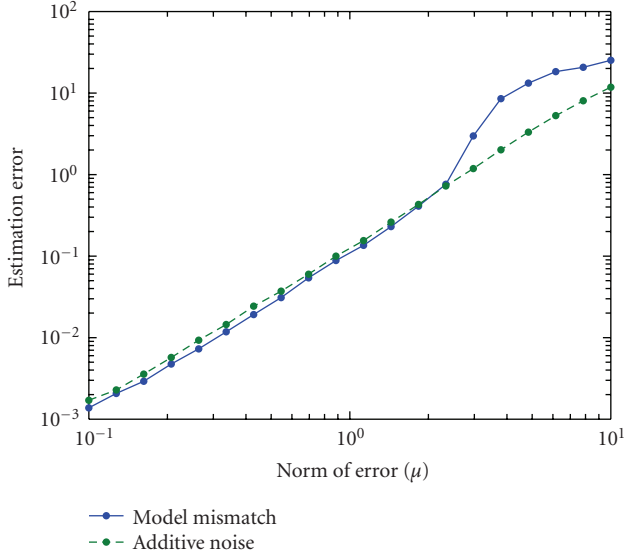


FIGURE 4: The graph shows the connection between the norm of the additive error  $\|\mathbf{N}\|_F$  and the estimation error of the underlying model  $J_{(\mathbf{W},\mathbf{H})}(\mathbf{W}', \mathbf{H}')$ . The two noise matrices from Example 4,  $\mathbf{N}_N$  and  $\mathbf{N}_M$ , are plotted. In this example, the curves are aligned for small errors, and for larger errors, the model error  $\mathbf{N}_M$  results in much larger estimation errors.

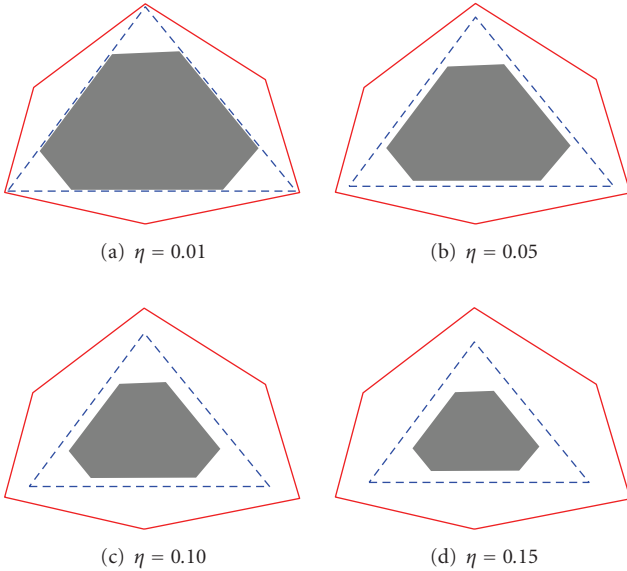


FIGURE 5: Data constructed as in Example 5 and plotted in the same manner as in Figure 1, that is, the dashed triangle is the desired solution, the solid line is the border of the dual of  $\mathbf{H}$ , and the shaded area is the positive span of  $\mathbf{W}$ . In all the plots,  $\alpha$  equals 0.3 and  $\eta$  equals 0.01, 0.05, 0.1, and 0.15.

sample space (the possible outcome) of a random variable that lead to unique NMF will be investigated.

**Theorem 7.** Let the row vectors of  $\mathbf{W}$  be generated by the random variable  $\mathbf{X}_W$  and let the column vectors of  $\mathbf{H}$  be generated by a random variable  $\mathbf{X}_H$ . If the sample space of

$\mathbf{X}_W$  is strongly boundary close and the sample space of  $\mathbf{X}_H$  is sufficiently spread, then for all  $\epsilon > 0$  and  $k < 1$ , there exist  $N_\epsilon$  and  $M_\epsilon$  such that

$$p\left(\min_{\mathbf{D}, \mathbf{P}}(\|\mathbf{D}\mathbf{P}\mathbf{Q} - \mathbf{I}\|_F) < \epsilon\right) > k, \quad (22)$$

where  $\mathbf{Q}$  is any matrix such that  $\mathbf{W}\mathbf{Q}$  and  $\mathbf{Q}^{-1}\mathbf{H}$  are nonnegative and the data size  $\mathbf{R} \in \mathbb{R}_+^{n \times m}$  is such that  $n > N_\epsilon$  and  $m > M_\epsilon$ .

*Proof.* If the data is scaled,  $\mathbf{D}_1\mathbf{R}\mathbf{D}_2$ , it does not change the nonuniqueness of the solutions when measured by the  $\mathbf{Q}$  matrix. The proof is therefore done on the normalized versions of  $\mathbf{W}$  and  $\mathbf{H}$ . Let  $\mathbf{Y}_W$  and  $\mathbf{Y}_H$  be the normalized version of  $\mathbf{X}_W$  and  $\mathbf{X}_H$ . There exist finite sets  $\bar{\mathbf{W}}$  and  $\bar{\mathbf{H}}$  of vectors in the closure of  $\mathbf{Y}_W$  and  $\mathbf{Y}_H$  that are strongly boundary close and sufficiently spread. By Theorem 5, it is known that  $\bar{\mathbf{V}} = \bar{\mathbf{W}}\bar{\mathbf{H}}$  is unique. By increasing the number of vectors sampled from  $\mathbf{Y}_W$  and  $\mathbf{Y}_H$ , for any  $\epsilon' > 0$ , there will be two subsets of the vectors,  $\mathbf{W}'$  and  $\mathbf{H}'$ , that with a probability larger than any  $k < 1$  will fulfil

$$\epsilon' > \|\bar{\mathbf{W}} - \mathbf{W}'\|_F + \|\bar{\mathbf{H}} - \mathbf{H}'\|_F. \quad (23)$$

It is possible to use Corollary 1 on this subset. The fact that limiting  $\min_{\mathbf{D}, \mathbf{P}}(\|\mathbf{D}\mathbf{P}\mathbf{Q} - \mathbf{I}\|_F)$  is equivalent to limiting (21) when the vectors are normalized concludes the proof.  $\square$

*Example 6.* Let all the elements in  $\mathbf{H}$  be exponential i.i.d. and therefore generated with a sufficiently spread sample space. Additionally, let each row in  $\mathbf{W}$  be exponential i.i.d. plus a random vector with the sample space  $\left\{\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\}$  and thereby strongly boundary close. In Figure 6, the above variables are shown for the following four matrix sizes  $\mathbf{R} \in \{\mathbb{R}^{10 \times 10}, \mathbb{R}^{40 \times 40}, \mathbb{R}^{100 \times 100}, \mathbb{R}^{500 \times 500}\}$ . This completes the example.

## 8. Discussion

The approach in this paper is to investigate when nonnegativity leads to uniqueness in connection with NMF,  $\mathbf{V} \approx \mathbf{R} = \mathbf{W}\mathbf{H}$ . Nonnegativity is the only assumption for the theorems, and the theorems therefore cannot be used as argument for an NMF to be nonunique if there is additional information about  $\mathbf{W}$  or  $\mathbf{H}$ . An example with stronger uniqueness results is the sparse NMF algorithm of Hoyer [12] built on the assumption that the row vectors in  $\mathbf{H}$  have known ratios between the  $L_1$  norm and the  $L_2$  norm. Theis et al. [13] have investigated uniqueness in this situation and shown strong uniqueness results. Another example is data matrices with an added constant on each row. For this situation, the affine NMF algorithm [14] can make NMF unique even though the setup violates Theorem 3 in this paper.

As shown in Figure 4, the type of noise greatly influences on the error curves. In applications where noise is introduced because the additive model does not hold as, for example, when  $\mathbf{V}$  is pictures or spectra, it is possible to influence the noise by making a nonlinear function on the elements



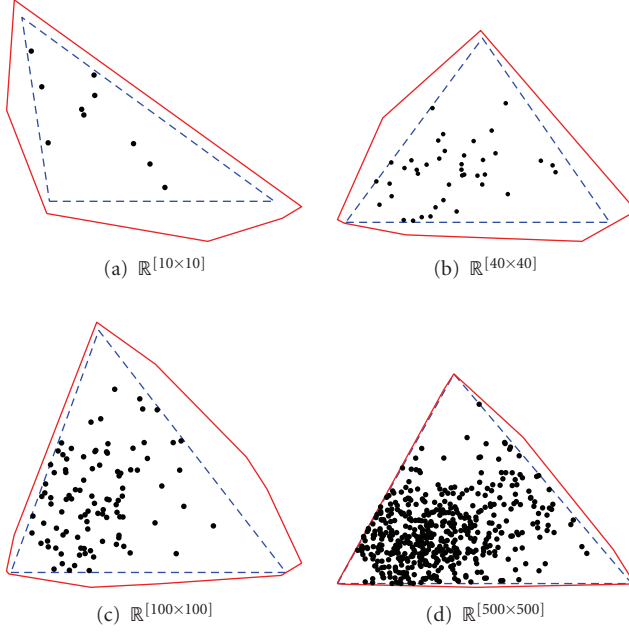


FIGURE 6: The figure shows data constructed as in Example 6 plotted in the same manner as the previous figure with the exception that each row vector of  $\mathbf{W}$  is plotted instead of the positive span of the vectors. The size of  $\mathbf{R}$  is shown under each plot.

of  $\mathbf{V}$ . Such a nonlinear function is introduced in [15] and experiments show that it improves the results. A theoretical framework for finding good nonlinear functions will be interesting to investigate.

The sufficiently spread condition defined in Section 5 has an important role for unique NMF due to Lemma 3. The sufficiently spread assumption is seen indirectly in related areas where it also leads to unique solutions, for example, in [7] where the groundedness assumption leads to variables with a sufficiently spread sample space. If the matrix  $\mathbf{H}$  is sufficiently spread, then the columns in  $\mathbf{W}$  will occur (almost) alone as columns in  $\mathbf{V}$ . Deville [16] uses the “occur alone” assumption, and thereby sufficiently spread assumption, to make blind source separation possible.

## 9. Conclusion

We have investigated the uniqueness of NMF from three different viewpoints as follows:

- (i) uniqueness in noise free situations;
- (ii) the estimation error of the underlying model when a matrix with unique NMF is added with noise; and
- (iii) the random processes that lead to matrices where the underlying model can be estimated with small errors.

By doing this, we have shown that it is possible to make many novel and useful characterizations that can be used as theoretical underpinning for using the numerous NMF algorithms. Several open issues can be found in all the three viewpoints that, if addressed, will give a better understanding of nonnegative matrix factorization.

## APPENDIX

*Proof of Theorem 4.* The theorem state that  $\mathbf{W} = \mathbf{W}\mathbf{I}$  is a unique NMF. To proof this, it is shown that the condition for Theorem 1 is fulfilled. The positive orthant is self-dual ( $\mathbf{I} = \mathbf{I}^{-1}$ ) and thereby  $\mathcal{Q} \subseteq \mathcal{P}$ , where  $\mathcal{Q}$  is an  $r$ -order simplicial cone that contains  $\text{span}^+(\mathbf{W}^T)$ . Let the set of row vectors in  $\mathbf{W}$  be denoted by  $\mathcal{W}$ . An  $r$ -order simplicial cone, like  $\mathcal{Q}$ , is a closed set and it therefore needs to contain the closure of  $\mathcal{W}$  denoted by  $\overline{\mathcal{W}}$ . The two items in Definition 6 of strongly boundary close can be reformulated for  $\overline{\mathcal{W}}$  that contains the border:

- (1)  $\mathbf{s}_n^j = 0$  for all  $j$ ,
- (2) the vectors  $[\mathbf{b}^1, \dots, \mathbf{b}^{d-n}]$  are linearly independent.

The rest of the proof follows by induction. If  $r = 2$ , then  $\overline{\mathcal{W}} = \mathcal{P}$  and is therefore unique. Let therefore  $r > 2$ . Then  $r - 1$  linearly independent vectors in  $\overline{\mathcal{W}}$  have zero as the first element, and  $r - 1$  of the basis vectors therefore need to have zero in the first element. In other words, there is only one basis vector with a nonzero first element. Let us call this vector  $\mathbf{b}^1$ . For all  $j > 1$  there is a vector in  $\overline{\mathcal{W}}$  which is nonnegative in the first element and zero in the  $j$ th element, so all the elements in  $\mathbf{b}^1$  except the first have to be zero. The proof is completed by seeing that if the first element is removed from the vectors in  $\overline{\mathcal{W}}$ , it is still strongly boundary close and the problem is therefore the  $r - 1$  dimensional problem.  $\square$

*Proof of Theorem 6.* Let  $\mathcal{G}$  be the open set of all  $\mathbf{W}'$ ,  $\mathbf{H}'$  pairs that are close to  $\mathbf{W}$  and  $\mathbf{H}$ ,

$$\mathcal{G} = \{[\mathbf{W}', \mathbf{H}'] \mid J_{(\mathbf{W}, \mathbf{H})}(\mathbf{W}', \mathbf{H}') < \epsilon\}. \quad (\text{A.1})$$

Let  $\overline{\mathcal{G}}$  be the set of all nonnegative  $\check{\mathbf{W}}$ ,  $\check{\mathbf{H}}$  pairs that are not in  $\mathcal{G}$  and where  $\max(\check{\mathbf{W}}, \check{\mathbf{H}}) \leq \sqrt{1 + \max(\mathbf{R})}$ . The uniqueness of  $\mathbf{R}$  ensures that

$$\|\mathbf{R} - \check{\mathbf{W}}\check{\mathbf{H}}\|_F > 0, \quad (\text{A.2})$$

for all  $[\check{\mathbf{W}}, \check{\mathbf{H}}] \in \overline{\mathcal{G}}$ . The fact that the Frobenius norm is continuous,  $\overline{\mathcal{G}}$  is a closed bounded set, and the statement above is positive ensures that

$$\min_{[\check{\mathbf{W}}, \check{\mathbf{H}}] \in \overline{\mathcal{G}}} (\|\mathbf{R} - \check{\mathbf{W}}\check{\mathbf{H}}\|_F) = \delta' > 0, \quad (\text{A.3})$$

since a continuous function attains its limits on a closed bounded set [17, Theorem 4.28]. The  $\check{\mathbf{W}}$ ,  $\check{\mathbf{H}}$  pairs that are not in  $\mathcal{G}$  and where  $\max(\check{\mathbf{W}}, \check{\mathbf{H}}) > \sqrt{1 + \max(\mathbf{R})}$  can either be transformed by a diagonal matrix into a matrix pair from  $\overline{\mathcal{G}}$ ,  $[\check{\mathbf{W}}\mathbf{D}, \mathbf{D}^{-1}\check{\mathbf{H}}] \in \overline{\mathcal{G}}$ , having the same product ( $\check{\mathbf{W}}\check{\mathbf{H}}$ ) or it can be transformed into a pair where both  $\check{\mathbf{W}}$  and  $\check{\mathbf{H}}$  have large elements, that is,

$$\max(\check{\mathbf{W}}\check{\mathbf{H}}) > \sqrt{1 + \max(\mathbf{R})}^2 = 1 + \max(\mathbf{R}), \quad (\text{A.4})$$

and thereby  $\|\mathbf{R} - \check{\mathbf{W}}\check{\mathbf{H}}\|_F > 1$ .

Select  $\delta$  to be  $\delta = \min(1, \delta')/2$ . The error of the desired solution  $\mathbf{R} = \mathbf{W}\mathbf{H}$  can be bounded by  $\|\mathbf{V} - \mathbf{R}\|_F = \|\mathbf{N}\|_F < \delta$ .

Let  $\tilde{\mathbf{V}}$  be any matrix constructed by a nonnegative matrix pair not from  $\mathcal{G}$ . Because of the way  $\delta$  is selected,  $\|\tilde{\mathbf{V}} - \mathbf{R}\|_F \geq 2\delta$ . By the triangle inequality, we get

$$\begin{aligned} \|\tilde{\mathbf{V}} - \mathbf{V}\|_F + \|\mathbf{V} - \mathbf{R}\|_F &\geq \|\tilde{\mathbf{V}} - \mathbf{R}\|_F \\ \|\tilde{\mathbf{V}} - \mathbf{V}\|_F &\geq \|\tilde{\mathbf{V}} - \mathbf{R}\|_F - \|\mathbf{V} - \mathbf{R}\|_F \\ &> 2\delta - \delta = \delta > \|\mathbf{V} - \mathbf{R}\|_F. \end{aligned} \quad (\text{A.5})$$

All solutions that are not in  $\mathcal{G}$  therefore have a larger error than  $\mathbf{WH}$  and will not be the minimizer of the error.  $\square$

## Acknowledgments

This research was supported by the Intelligent Sound project, Danish Technical Research Council Grant no. 26-02-0092. The work of M. G. Christensen is supported by the Parametric Audio Processing project, Danish Research Council for Technology, and Production Sciences Grant no. 274-06-0521. Part of this work was previously presented at a conference [18].

## References

- [1] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
- [2] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems 13*, pp. 556–562, MIT Press, Cambridge, Mass, USA, 2000.
- [3] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 155–173, 2007.
- [4] A. Berman, "Problem 73-14, rank factorization of nonnegative matrices," *SIAM Review*, vol. 15, no. 3, p. 655, 1973.
- [5] L. Thomas, "Solution to problem 73-14, rank factorizations of nonnegative matrices," *SIAM Review*, vol. 16, no. 3, pp. 393–394, 1974.
- [6] D. Donoho and V. Stodden, "When does non-negative matrix factorization give a correct decomposition into parts?" in *Advances in Neural Information Processing Systems 16*, pp. 1141–1148, MIT Press, Cambridge, Mass, USA, 2004.
- [7] M. Plumbley, "Conditions for nonnegative independent component analysis," *IEEE Signal Processing Letters*, vol. 9, no. 6, pp. 177–180, 2002.
- [8] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, USA, 1st edition, 1970.
- [9] H. Minc, *Nonnegative Matrices*, John Wiley & Sons, New York, NY, USA, 1st edition, 1988.
- [10] G. H. Golub and C. F. V. Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, Md, USA, 3rd edition, 1996.
- [11] P. Smaragdis and J. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA '03)*, pp. 177–180, New Paltz, NY, USA, October 2003.
- [12] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *The Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.
- [13] F. Theis, K. Stadlthanner, and T. Tanaka, "First results on uniqueness of sparse non-negative matrix factorization," in *Proceedings of the 13th European Signal Processing Conference (EUSIPCO '05)*, Antalya, Turkey, September 2005.
- [14] H. Laurberg and L. K. Hansen, "On affine non-negative matrix factorization," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, vol. 2, pp. 653–656, Honolulu, Hawaii, USA, April 2007.
- [15] M. N. Schmidt, J. Larsen, and F.-T. Hsiao, "Wind noise reduction using non-negative sparse coding," in *Proceedings of the IEEE Workshop on Machine Learning for Signal Processing (MLSP '07)*, pp. 431–436, Thessaloniki, Greece, August 2007.
- [16] Y. Deville, "Temporal and time-frequency correlation-based blind source separation methods," in *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA '03)*, pp. 1059–1064, Nara, Japan, April 2003.
- [17] T. M. Apostol, *Mathematical Analysis*, Addison-Wesley, Reading, Mass, USA, 2nd edition, 1974.
- [18] H. Laurberg, "Uniqueness of non-negative matrix factorization," in *Proceedings of the 14th IEEE/SP Workshop on Statistical Signal Processing (SSP '07)*, pp. 44–48, Madison, Wis, USA, August 2007.